Chapter 3 - Conditional Propability & Idependence

|*See examples of chapter 3 |*

3.1 - Introduction

- One of the most important concept in probability → conditional probability.

1) Calculating probabilities when some partial information concerning the result of an experiment is available; ⇝ desired probabilities are conditional

2) When no partial information is available, conditional probabilities can often be used to compute the desired probabilites.

3.2 - Conditional Probabilities

If $\mathbb{P}(F) > 0$, then

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(EF)}{\mathbb{P}(F)}$$

EXAMPLE 2d

A total of n balls are sequentially and randomly chosen, without replacement, from an urn containing r red and b blue balls ($n \leq r + b$). Given that k of the n balls are blue, what is the conditional probability that the first ball chosen is blue?

Solution. If we imagine that the balls are numbered, with the blue balls having numbers 1 through b and the red balls $b + 1$ through $b + r$, then the outcome of the experiment of selecting n balls without replacement is a vector of distinct integers $x_1, \ldots, x_n$, where each x_i is between 1 and $r + b$. Moreover, each such vector is equally likely to be the outcome. So, given that the vector contains k blue balls (that is, it contains k values between 1 and b), it follows that each of these outcomes is equally likely. But because the first ball chosen is, therefore, equally likely to be any of the n chosen balls, of which k are blue, it follows that the desired probability is k/n.

If we did not choose to work with the reduced sample space, we could have solved the problem by letting B be the event that the first ball chosen is blue and B_k be the event that a total of k blue balls are chosen. Then

$$\begin{aligned} P(B|B_k) &= \frac{P(BB_k)}{P(B_k)} \\ &= \frac{P(B_k|B)P(B)}{P(B_k)} \end{aligned}$$

Now, $P(B_k|B)$ is the probability that a random choice of $n - 1$ balls from an urn containing r red and $b - 1$ blue balls results in a total of $k - 1$ blue balls being chosen; consequently,

$$P(B_k|B) = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}}$$

Using the preceding formula along with

$$P(B) = \frac{b}{r + b}$$

and the hypergeometric probability

$$P(B_k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{r+b}{n}}$$

again yields the result that

$$P(B|B_k) = \frac{k}{n}$$ ∎

Multiplying both sides of Equation (2.1) by $P(F)$, we obtain

$$P(EF) = P(F)P(E|F) \tag{2.2}$$

In words, Equation (2.2) states that the probability that both E and F occur is equal to the probability that F occurs multiplied by the conditional probability of E given that F occurred. Equation (2.2) is often quite useful in computing the probability of the intersection of events.

* Multiplication rule:

The multiplication rule

$$P(E_1E_2E_3\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

To prove the multiplication rule, just apply the definition of conditional probability to its right-hand side, giving

$$P(E_1)\frac{P(E_1E_2)}{P(E_1)}\frac{P(E_1E_2E_3)}{P(E_1E_2)}\cdots\frac{P(E_1E_2\cdots E_n)}{P(E_1E_2\cdots E_{n-1})} = P(E_1E_2\cdots E_n)$$

3.3 - Bayes's Formula

Let E & F be events.

We may express E as

$$E = EF \cup EF^c$$

⇒ In order for an outcome to be in E, it must either be in both E & F, or be in E but not F

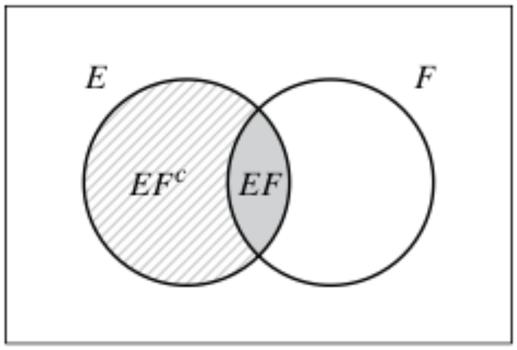

FIGURE 3.1: $E = EF \cup EF^c$. EF = Shaded Area; EF^c = Striped Area

As we can see EF & EF^c are clearly mutually exclusive, ∴ by Axiom 3

$$\boxed{\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E|F)P(F) + P(E|F^c)P(F^c) \\ &= P(E|F)P(F) + P(E|F^c)[1 - P(F)] \end{aligned}}$$

Equation above states.
That the probability of the event E is a weighted average of the conditional probability of E given that F has occured. and the conditional probabil

of E given that F has not occured.

=> Extremely useful formula, because its use often enables us to determine the probability of an event by first "conditioning" upon whether or some second event has occured.

EXAMPLES:

Urn 1 initially has n red molecules and urn 2 has n blue molecules. Molecules are randomly removed from urn 1 in the following manner.

- After each removal from urn 1, a molecule is taken from urn 2 (if urn 2 has any molecules) and placed in urn 1

The process continues until all the

molecules have been removed.
(Thus, there are $2n$ removals in all)
a) Find $P(R)$, where R is the event that the final molecule removed from urn 1 is red.
b) Repeat the problem when urn 1 initially has r_1 red molecules & b_1 blue molecules and urn 2 has initially r_2 red molecules & b_2 blue molecules.

Solution. (a) Focus attention on any particular red molecule, and let F be the event that this molecule is the final one selected. Now, in order for F to occur, the molecule in question must still be in the urn after the first n molecules have been removed (at which time urn 2 is empty). So, letting N_i be the event that this molecule is not the ith molecule to be removed, we have

$$\begin{aligned} P(F) &= P(N_1 \cdots N_n F) \\ &= P(N_1)P(N_2|N_1)\cdots P(N_n|N_1\cdots N_{n-1})P(F|N_1\cdots N_n) \\ &= \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{1}{n}\right)\frac{1}{n} \end{aligned}$$

where the preceding formula uses the fact that the conditional probability that the molecule under consideration is the final molecule to be removed, given that it is still in urn 1 when only n molecules remain, is, by symmetry, $1/n$.

Therefore, if we number the n red molecules and let R_j be the event that red molecule number j is the final molecule removed, then it follows from the preceding formula that

$$P(R_j) = \left(1-\frac{1}{n}\right)^n \frac{1}{n}$$

Because the events R_j are mutually exclusive, we obtain

$$P(R) = P\left(\bigcup_{j=1}^{n} R_j\right) = \sum_{j=1}^{n} P(R_j) = (1-\frac{1}{n})^n \approx e^{-1}$$

(b) Suppose now that urn i initially has r_i red and b_i blue molecules, for $i = 1, 2$. To find $P(R)$, the probability that the final molecule removed is red, focus attention on any molecule that is initially in urn 1. As in part (a), it follows that the probability that this molecule is the final one removed is

$$p = \left(1-\frac{1}{r_1+b_1}\right)^{r_2+b_2} \frac{1}{r_1+b_1}$$

That is, $\left(1-\frac{1}{r_1+b_1}\right)^{r_2+b_2}$ is the probability that the molecule under consideration is still in urn 1 when urn 2 becomes empty, and $\frac{1}{r_1+b_1}$ is the conditional probability, given the preceding event, that the molecule under consideration is the final molecule removed. Hence, if we let O be the event that the last molecule removed is one of the molecules originally in urn 1, then

$$P(O) = (r_1+b_1)p = \left(1-\frac{1}{r_1+b_1}\right)^{r_2+b_2}$$

To determine $P(R)$, we condition on whether O occurs, to obtain

$$\begin{aligned} P(R) &= P(R|O)P(O) + P(R|O^c)P(O^c) \\ &= \frac{r_1}{r_1+b_1}\left(1-\frac{1}{r_1+b_1}\right)^{r_2+b_2} + \frac{r_2}{r_2+b_2}\left(1-\left(1-\frac{1}{r_1+b_1}\right)^{r_2+b_2}\right) \end{aligned}$$

If $r_1 + b_1 = r_2 + b_2 = n$, so that both urns initially have n molecules, then, when n is large,

$$P(L) \approx \frac{r_1}{r_1 + b_1} e^{-1} + \frac{r_2}{r_2 + b_2}(1 - e^{-1}) \qquad \blacksquare$$

The change in the probability of a hypothesis when new evidence is introduced can be expressed compactly in terms of the change in the *odds* of that hypothesis, where the concept of odds is defined as follows.

Definition

The odds of an event A are defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

That is, the odds of an event A tell how much more likely it is that the event A occurs than it is that it does not occur. For instance, if $P(A) = \frac{2}{3}$, then $P(A) = 2P(A^c)$, so the odds are 2. If the odds are equal to α, then it is common to say that the odds are "α to 1" in favor of the hypothesis.

Consider now a hypothesis H that is true with probability $P(H)$, and suppose that new evidence E is introduced. Then the conditional probabilities, given the evidence E, that H is true and that H is not true are respectively given by

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \qquad P(H^c|E) = \frac{P(E|H^c)P(H^c)}{P(E)}$$

Therefore, the new odds after the evidence E has been introduced are

$$\boxed{\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)} \frac{P(E|H)}{P(E|H^c)}} \tag{3.3}$$

That is, the new value of the odds of H is the old value, multiplied by the ratio of the conditional probability of the new evidence given that H is true to the conditional probability given that H is not true. Thus, Equation (3.3) verifies the result of Example 3f, since the odds, and thus the probability of H, increase whenever the new evidence is more likely when H is true than when it is false. Similarly, the odds decrease whenever the new evidence is more likely when H is false than when it is true.

3.4 - Independent Events

The previous examples of this chapter show that $P(E|F) \neq P(E)$ [unconditional probability of E]

In other words, knowing that F has occured generally changes the chances of E's occurance.

→ Special case:

where $P(E|F)$ does in fact equal $P(E)$, we say that E is <u>independent</u> of F.

⇝ E is independent of F if knowledge that F has occured does not change the probability that E occurs.

$$\Rightarrow P(E|F) = \frac{P(EF)}{P(F)}$$

↳ E is independent of F if $P(EF) = P(E)P(F)$

→ Symmetric in E & F shows that whenever E is independent of F, F is also independent of E.

EXAMPLE 4d

If we let E denote the event that the next president is a Republican and F the event that there will be a major earthquake within the next year, then most people would probably be willing to assume that E and F are independent. However, there would probably be some controversy over whether it is reasonable to assume that E is independent of G, where G is the event that there will be a recession within two years after the election. ■

We now show that if E is independent of F, then E is also independent of F^c.

Proposition 4.1. If E and F are independent, then so are E and F^c.

Proof. Assume that E and F are independent. Since $E = EF \cup EF^c$ and EF and EF^c are obviously mutually exclusive, we have

$$\begin{aligned} P(E) &= P(EF) + P(EF^c) \\ &= P(E)P(F) + P(EF^c) \end{aligned}$$

or, equivalently,

$$\begin{aligned} P(EF^c) &= P(E)[1 - P(F)] \\ &= P(E)P(F^c) \end{aligned}$$

and the result is proved. □

Thus, if E is independent of F, then the probability of E's occurrence is unchanged by information as to whether or not F has occurred.

Suppose now that E is independent of F and is also independent of G. Is E then necessarily independent of FG? The answer, somewhat surprisingly, is no, as the following example demonstrates.

3.5 - $\mathbb{P}(\cdot | F)$ is a Probability

- Conditional probabilities satisfy all properties of ordinary probabilities as is proved by Proposition 5.1 which shows that $\mathbb{P}(E|F)$ satisfies the 3 axioms of probability

Proposition 5.1.

(a) $0 \leq P(E|F) \leq 1$.
(b) $P(S|F) = 1$.
(c) If $E_i, i = 1, 2, \ldots$, *are mutually exclusive events, then*

$$P\left(\bigcup_1^\infty E_i|F\right) = \sum_1^\infty P(E_i|F)$$

Proof. To prove part (a), we must show that $0 \leq P(EF)/P(F) \leq 1$. The left-side inequality is obvious, whereas the right side follows because $EF \subset F$, which implies that $P(EF) \leq P(F)$. Part (b) follows because

$$P(S|F) = \frac{P(SF)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

Part (c) follows from

$$\begin{aligned}
P\left(\bigcup_{i=1}^\infty E_i|F\right) &= \frac{P\left(\left(\bigcup_{i=1}^\infty E_i\right)F\right)}{P(F)} \\
&= \frac{P\left(\bigcup_1^\infty E_iF\right)}{P(F)} \quad \text{since} \quad \left(\bigcup_1^\infty E_i\right)F = \bigcup_1^\infty E_iF \\
&= \frac{\sum_1^\infty P(E_iF)}{P(F)} \\
&= \sum_1^\infty P(E_i|F)
\end{aligned}$$

where the next-to-last equality follows because $E_iE_j = \emptyset$ implies that $E_iFE_jF = \emptyset$. □

---

†See N. Alon, J. Spencer, and P. Erdos, *The Probabilistic Method* (New York: John Wiley & Sons, Inc., 1992).

If we define $Q(E) = P(E|F)$, then, from Proposition 5.1, $Q(E)$ may be regarded as a probability function on the events of S. Hence, all of the propositions previously proved for probabilities apply to $Q(E)$. For instance, we have

$$Q(E_1 \cup E_2) = Q(E_1) + Q(E_2) - Q(E_1E_2)$$

or, equivalently,

$$P(E_1 \cup E_2|F) = P(E_1|F) + P(E_2|F) - P(E_1E_2|F)$$

Also, if we define the conditional probability $Q(E_1|E_2)$ by $Q(E_1|E_2) = Q(E_1E_2)/Q(E_2)$, then, from Equation (3.1), we have

$$Q(E_1) = Q(E_1|E_2)Q(E_2) + Q(E_1|E_2^c)Q(E_2^c) \tag{5.1}$$

Since

$$\begin{aligned} Q(E_1|E_2) &= \frac{Q(E_1E_2)}{Q(E_2)} \\ &= \frac{P(E_1E_2|F)}{P(E_2|F)} \\ &= \frac{\dfrac{P(E_1E_2F)}{P(F)}}{\dfrac{P(E_2F)}{P(F)}} \\ &= P(E_1|E_2F) \end{aligned}$$

Equation (5.1) is equivalent to

$$P(E_1|F) = P(E_1|E_2F)P(E_2|F) + P(E_1|E_2^cF)P(E_2^c|F)$$

SUMMARY

For events E and F, the conditional probability of E given that F has occurred is denoted by $P(E|F)$ and is defined by

$$P(E|F) = \frac{P(EF)}{P(F)}$$

The identity

$$P(E_1E_2\cdots E_n) = P(E_1)P(E_2|E_1)\cdots P(E_n|E_1\cdots E_{n-1})$$

is known as the *multiplication rule* of probability.

A valuable identity is

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$

which can be used to compute $P(E)$ by "conditioning" on whether F occurs.

$P(H)/P(H^c)$ is called the *odds* of the event H. The identity

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)\,P(E|H)}{P(H^c)P(E|H^c)}$$

shows that when new evidence E is obtained, the value of the odds of H becomes its old value multiplied by the ratio of the conditional probability of the new evidence when H is true to the conditional probability when H is not true.

Let F_i, $i = 1,\ldots,n$, be mutually exclusive events whose union is the entire sample space. The identity

$$P(F_j|E) = \frac{P(E|F_j)P(F_j)}{\displaystyle\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

is known as *Bayes's formula*. If the events F_i, $i = 1,\ldots,n$, are competing hypotheses, then Bayes's formula shows how to compute the conditional probabilities of these hypotheses when additional evidence E becomes available.

If $P(EF) = P(E)P(F)$, then we say that the events E and F are *independent*. This condition is equivalent to $P(E|F) = P(E)$ and to $P(F|E) = P(F)$. Thus, the events E and F are independent if knowledge of the occurrence of one of them does not affect the probability of the other.

The events $E_1,\ldots,E_n$ are said to be independent if, for any subset $E_{i_1},\ldots,E_{i_r}$ of them,

$$P(E_{i_1}\cdots E_{i_r}) = P(E_{i_1})\cdots P(E_{i_r})$$

For a fixed event F, $P(E|F)$ can be considered to be a probability function on the events E of the sample space.

* Important

Examples found in this chapter are very nice exercise